mri_dat_three_analysis

Philip Anderson; panders2@tamu.edu

6/17/2018

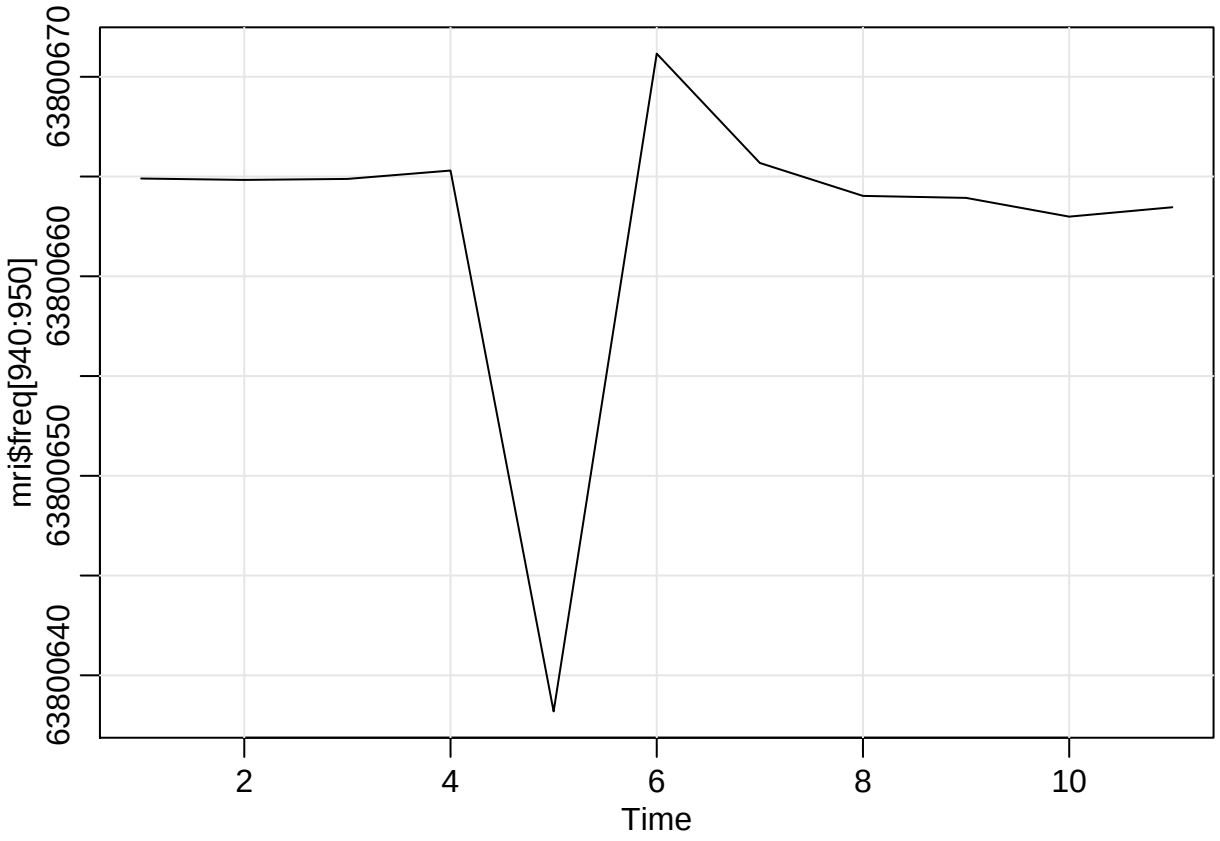
Objective

The purpose of this analysis is to try out some regressions of the data.

```
# read in the data
mri <- read.csv("/Users/panders2/Documents/schools/tamu/stat_626/project/stat_626_proj/mri_dat_one.csv")
# update field names
names(mri) <- c("hour", "minute", "freq", "int_pressure", "atm_pressure", "tot_pressure", "tesla")
```

First, replace the issue values with their imputations.

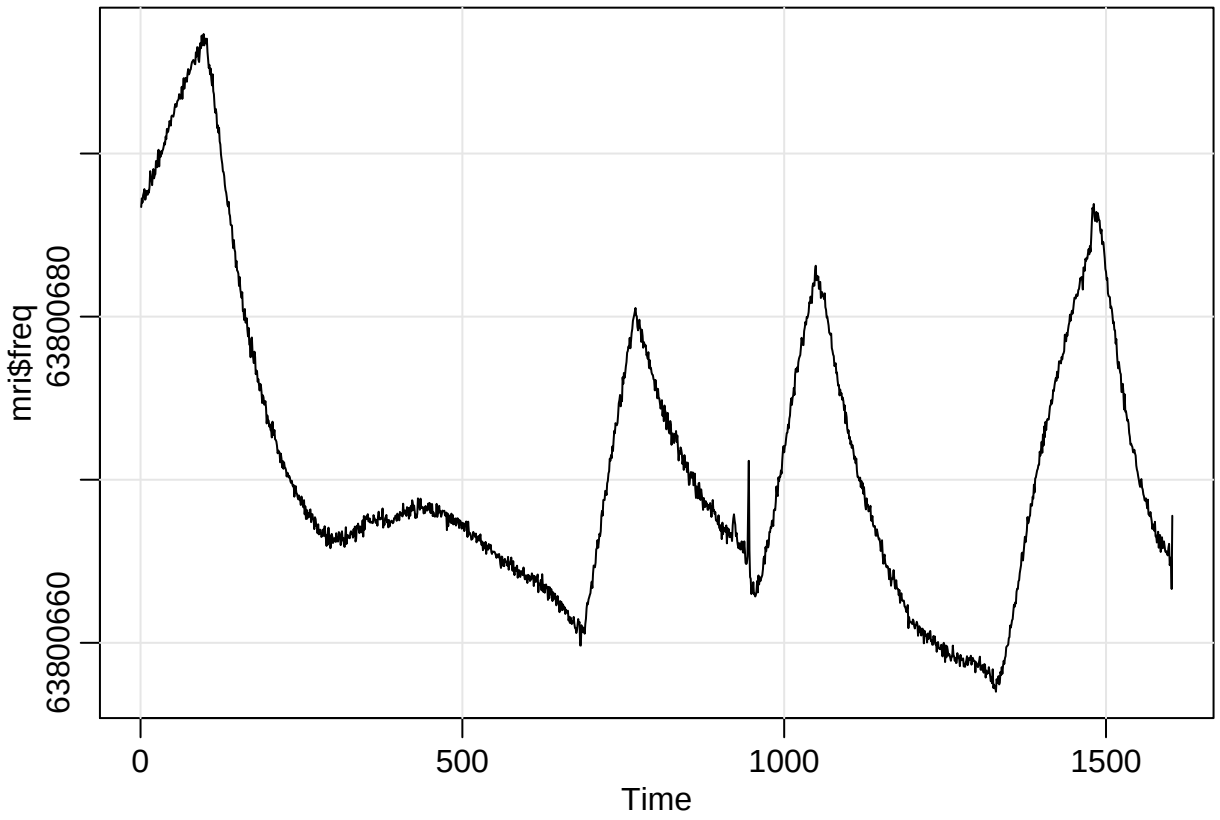
```
tsplot(mri$freq[940:950])
```



The problems are the points at index 945.

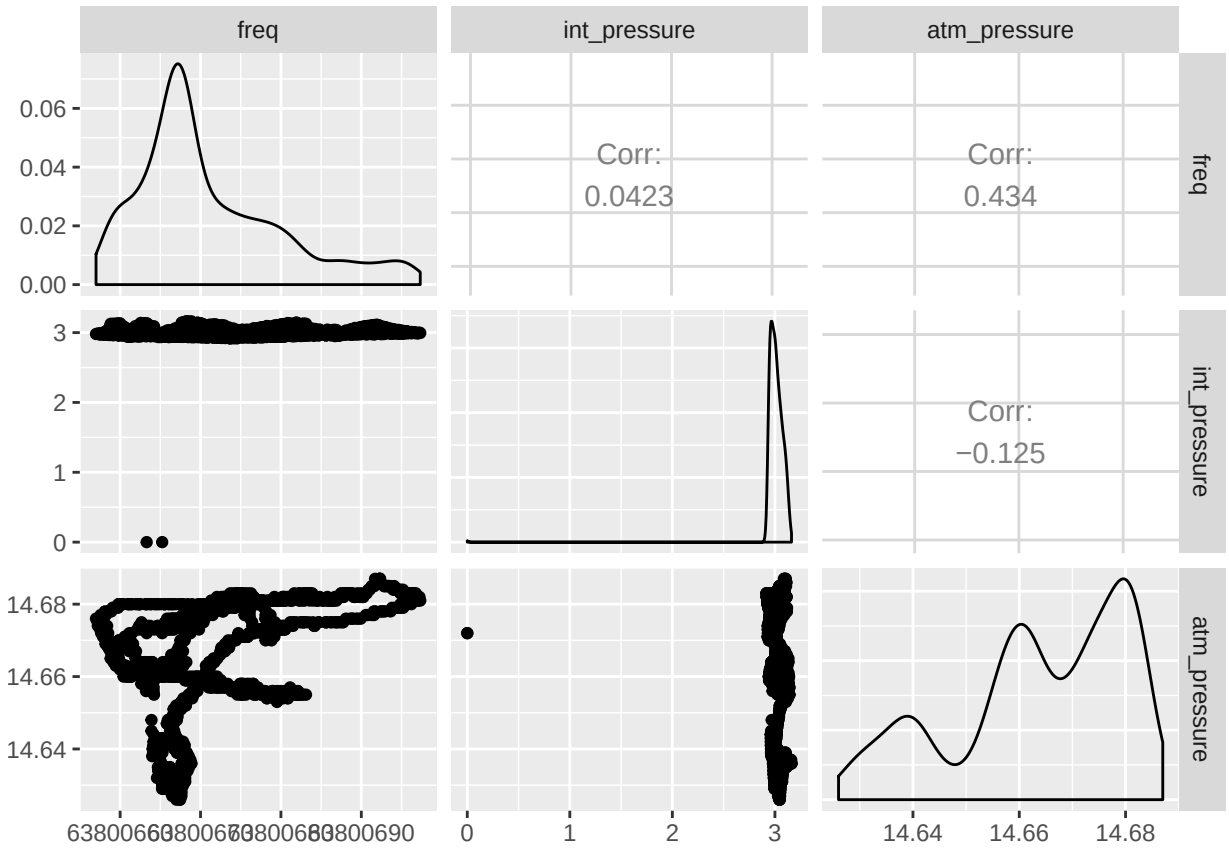
```
mri$freq[944] <- ((mri$freq[943] + mri$freq[945]) / 2)
```

```
astsa::tsplot(mri$freq)
```



That's better.

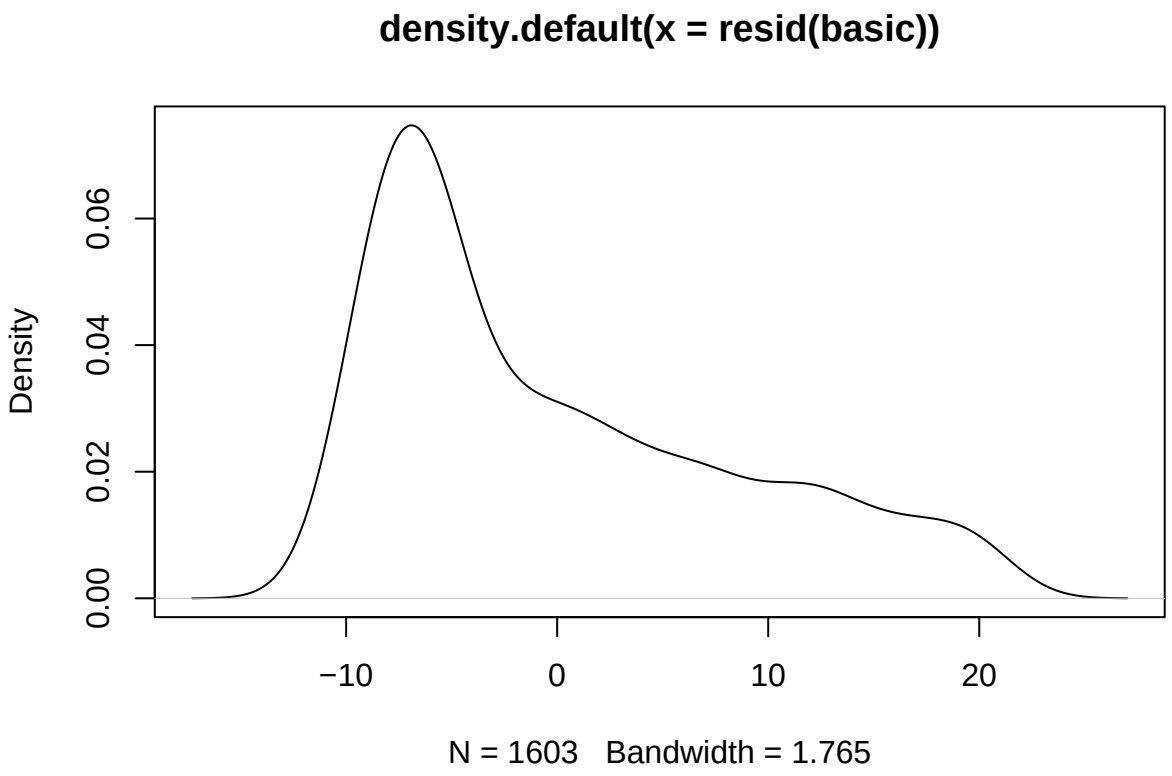
```
# parwise scatter plots of our data
GGally::ggpairs(mri%>%dplyr::select("freq", "int_pressure", "atm_pressure"))
```



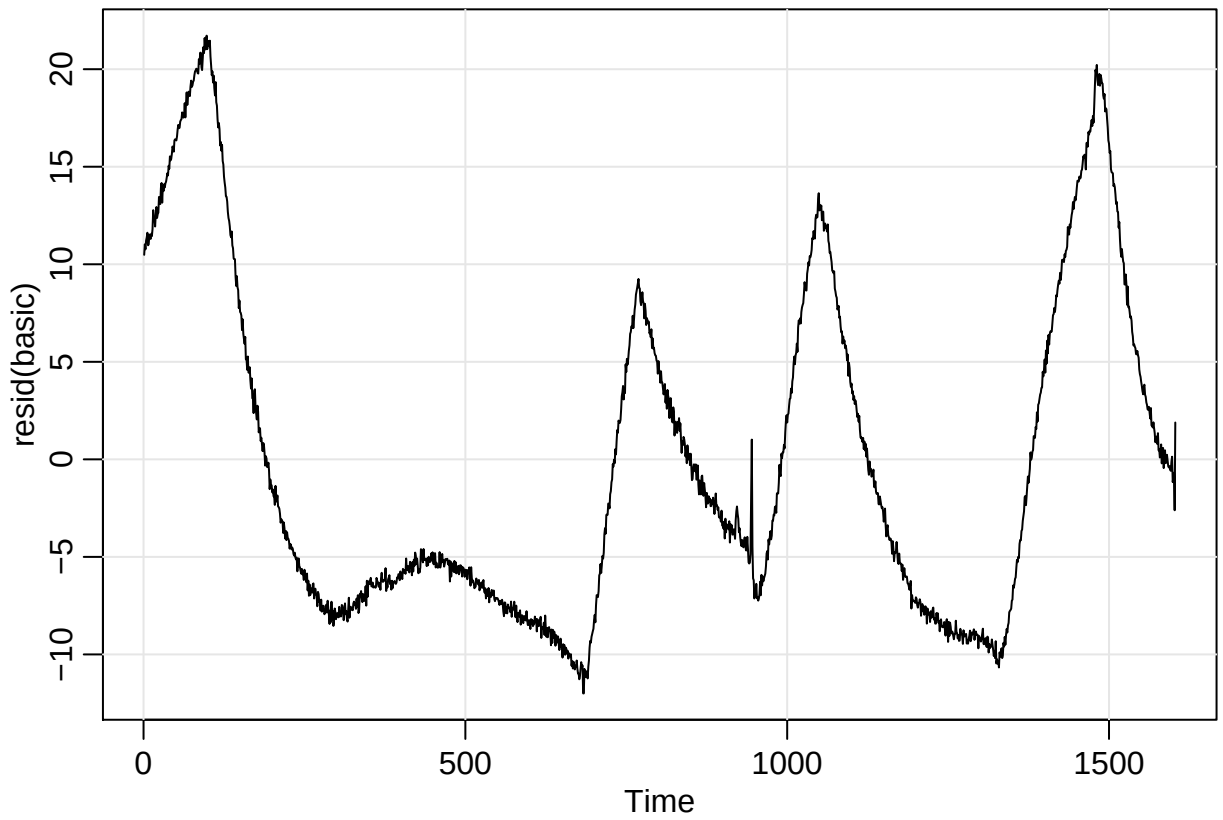
These are not particularly helpful, probably because of the time series nature of the data.

Fit naive linear regressions

```
basic <- lm(mri$freq ~ time(mri$freq))
plot(density(resid(basic)))
```



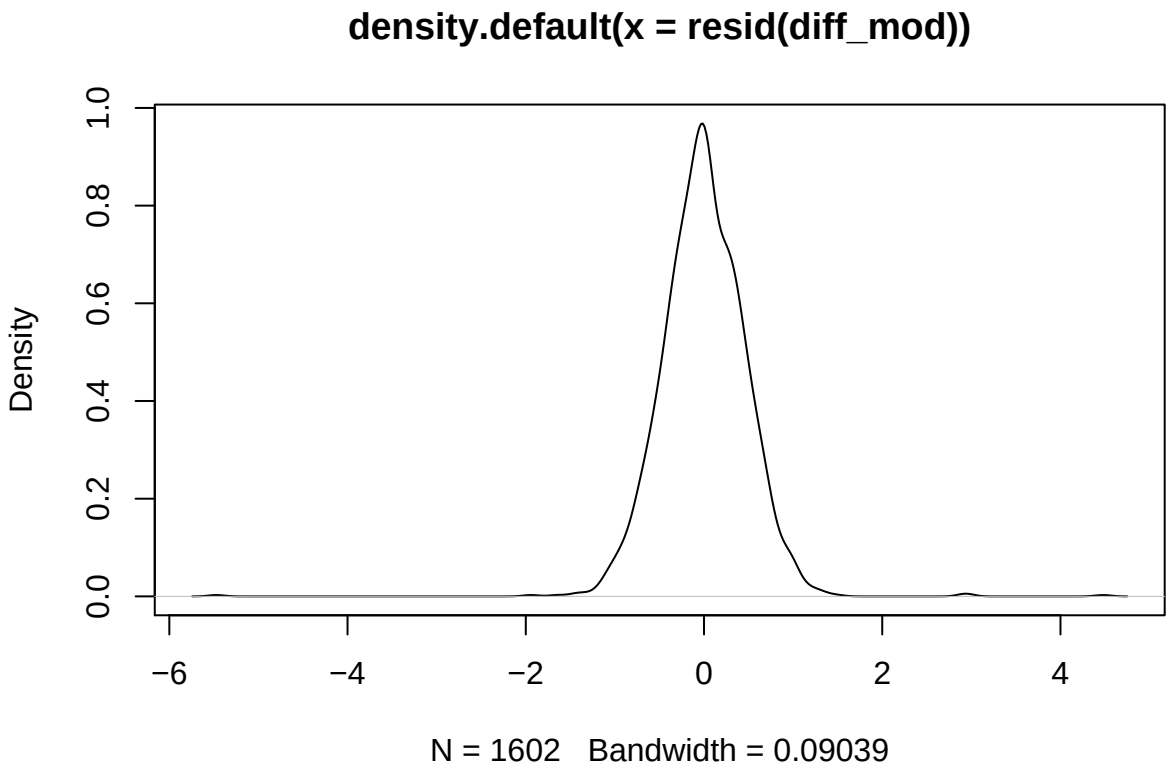

```
astsa::tsplot(resid(basic))
```



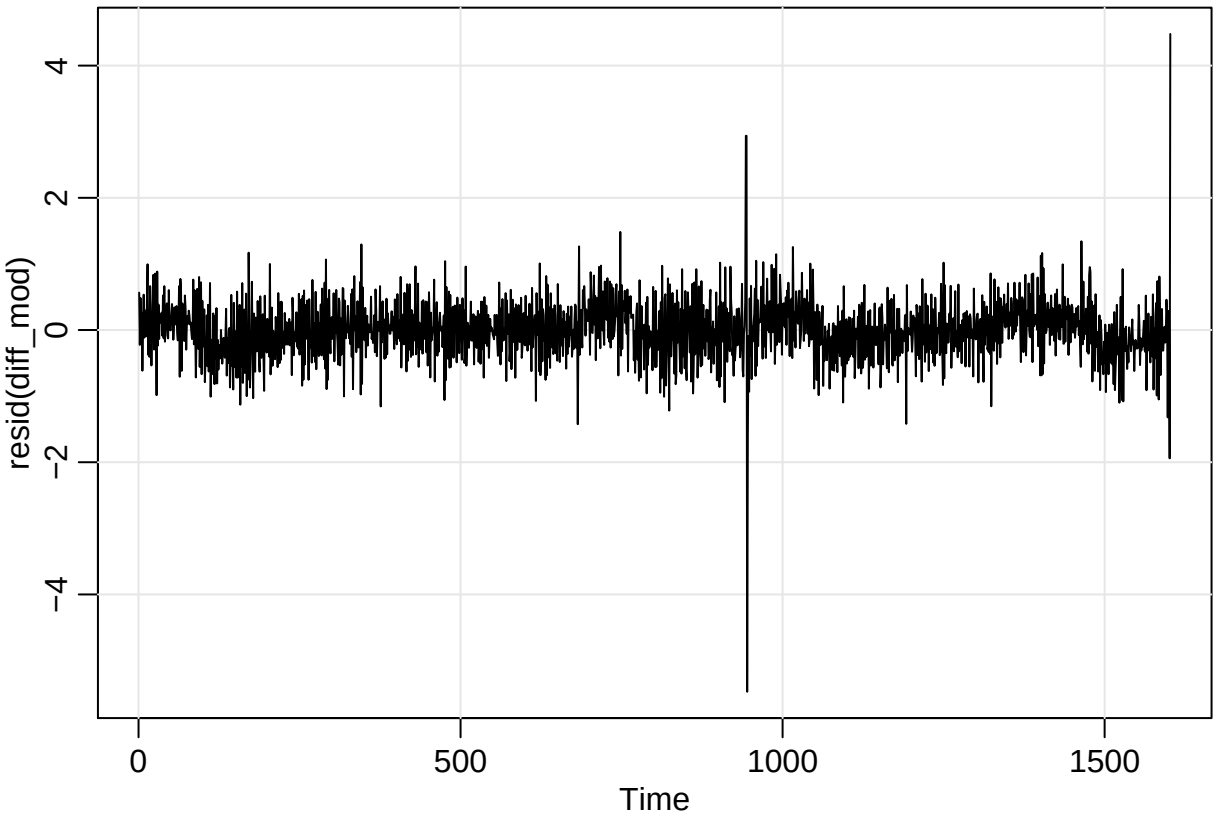
Errors are obviously serially correlated. Now try this with the differenced series to see if it makes a difference.

```r
diff_mod <- lm(diff(mri$freq) ~ time(mri$freq[2:length(mri$freq)]))

plot(density(resid(diff_mod)))
```




```r
astsa::tsplot(resid(diff_mod))
```

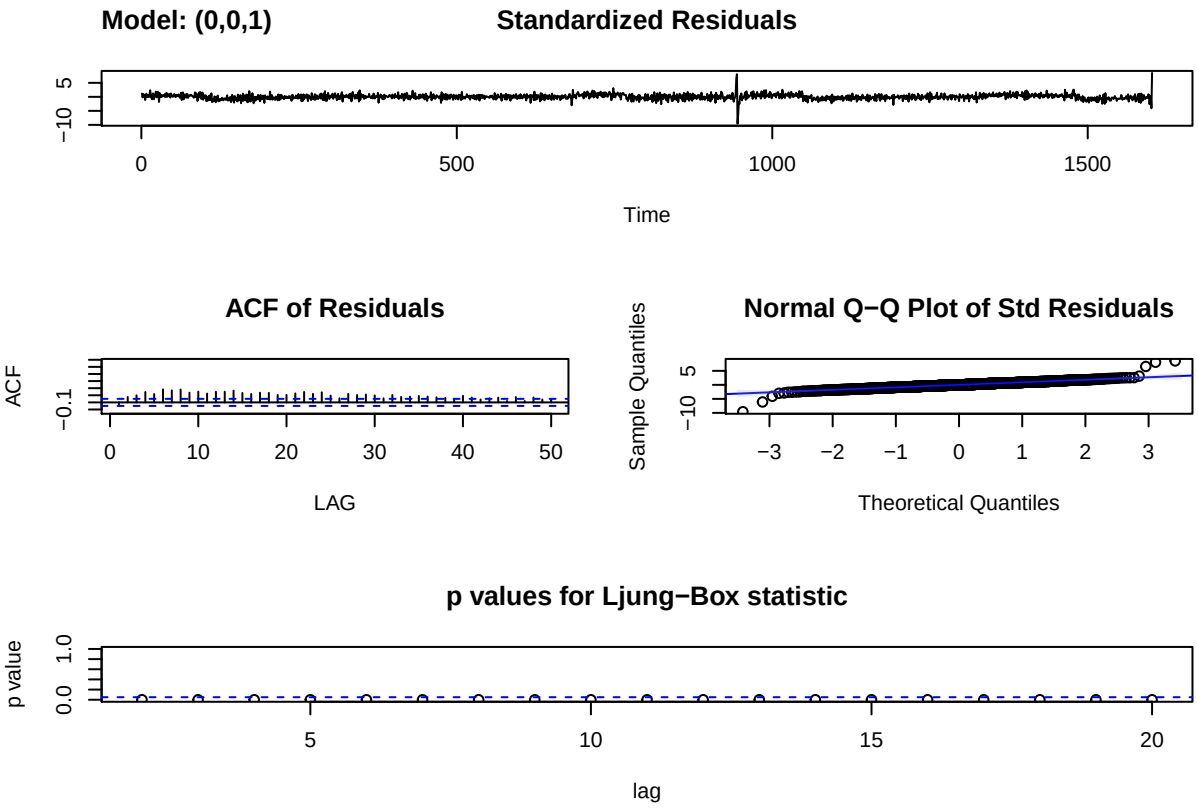


Looks better.

The ACF+PACF of the differenced series suggest an MA(1) model; let's try fitting this.

```
serial_mod <- sarima(
                     xdata=diff(mri$freq)
                     , p=c(0)
                     , d=c(0)
                     , q=c(1)
                     , xreg=c(time(mri$freq[2:length(mri$freq)]))
                     )
```

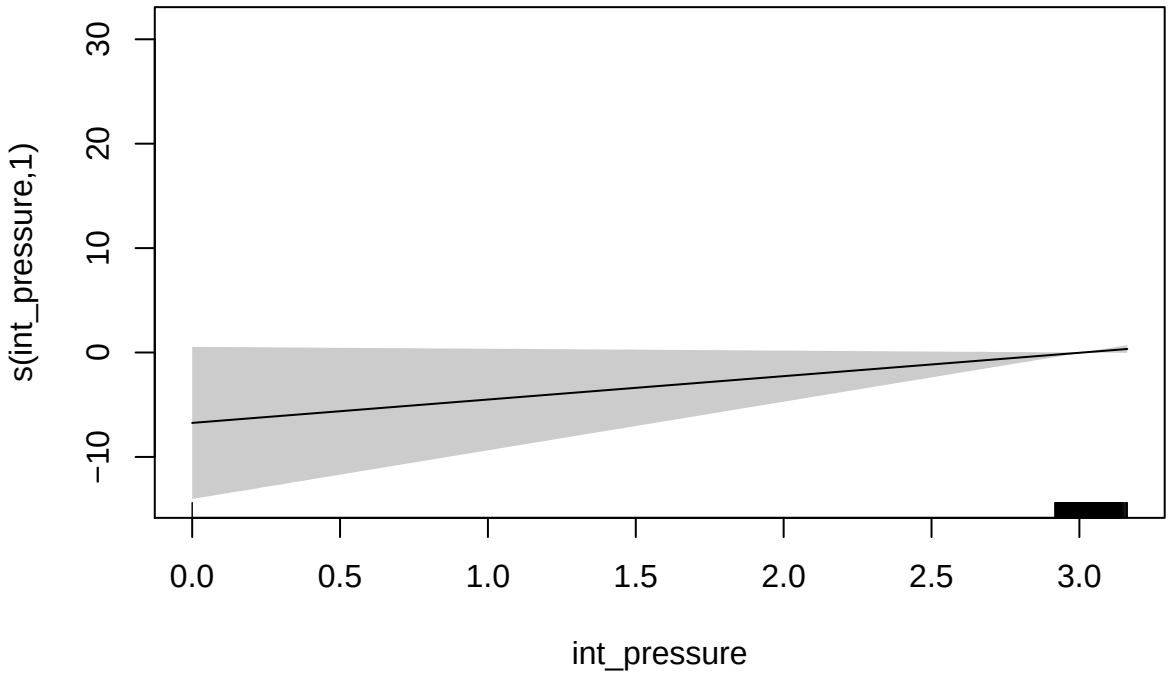
```
## initial  value -0.723974
## iter   2 value -0.764969
## iter   3 value -0.765137
## iter   4 value -0.765157
## iter   5 value -0.765158
## iter   6 value -0.765161
## iter   7 value -0.765161
## iter   8 value -0.765161
## iter   8 value -0.765161
## final  value -0.765161
## converged
## initial  value -0.765171
## iter   1 value -0.765171
## final  value -0.765171
## converged
```

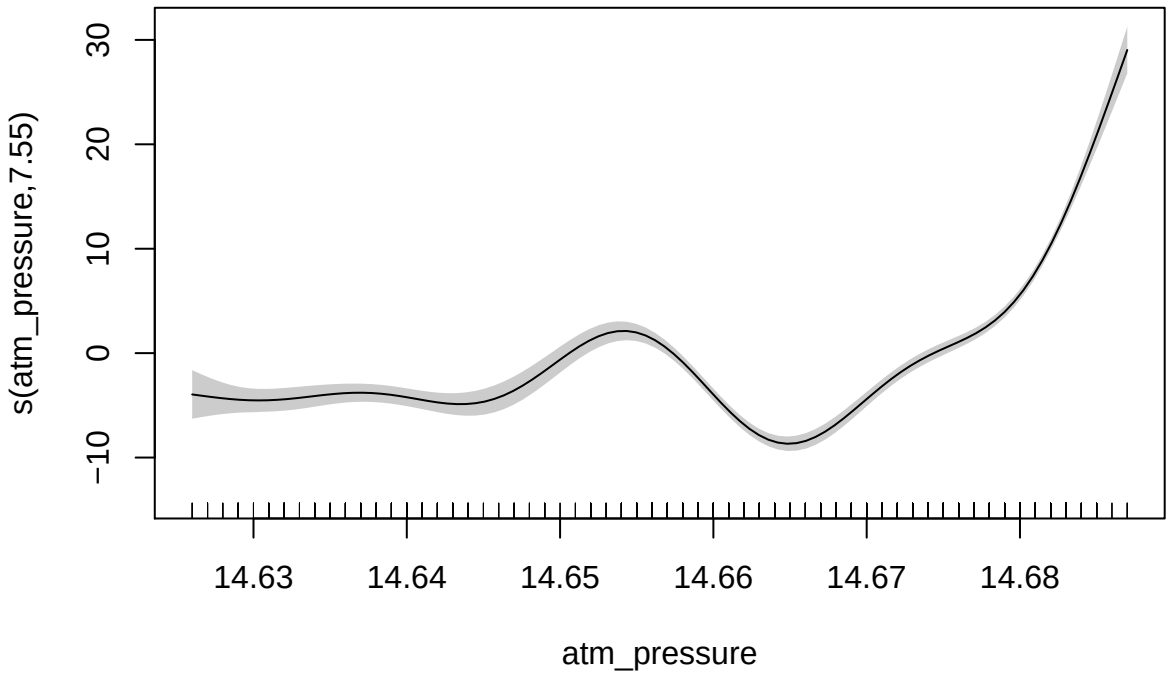


Looks like there are some problems.

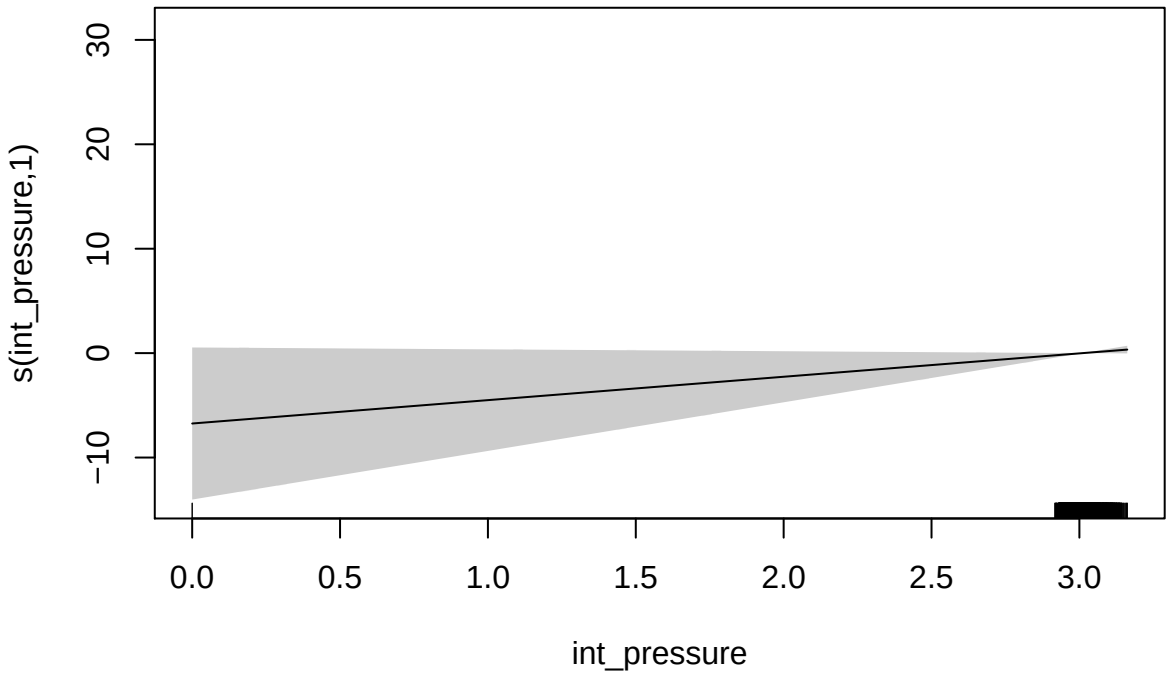
Try semi-parametric regression (Generalized Additive Model) to see if we can get some explanation for the strange-looking scatterplot. I'm not sure if there is a specialized version of these for time series, and we haven't covered them, but they might be somewhat helpful.

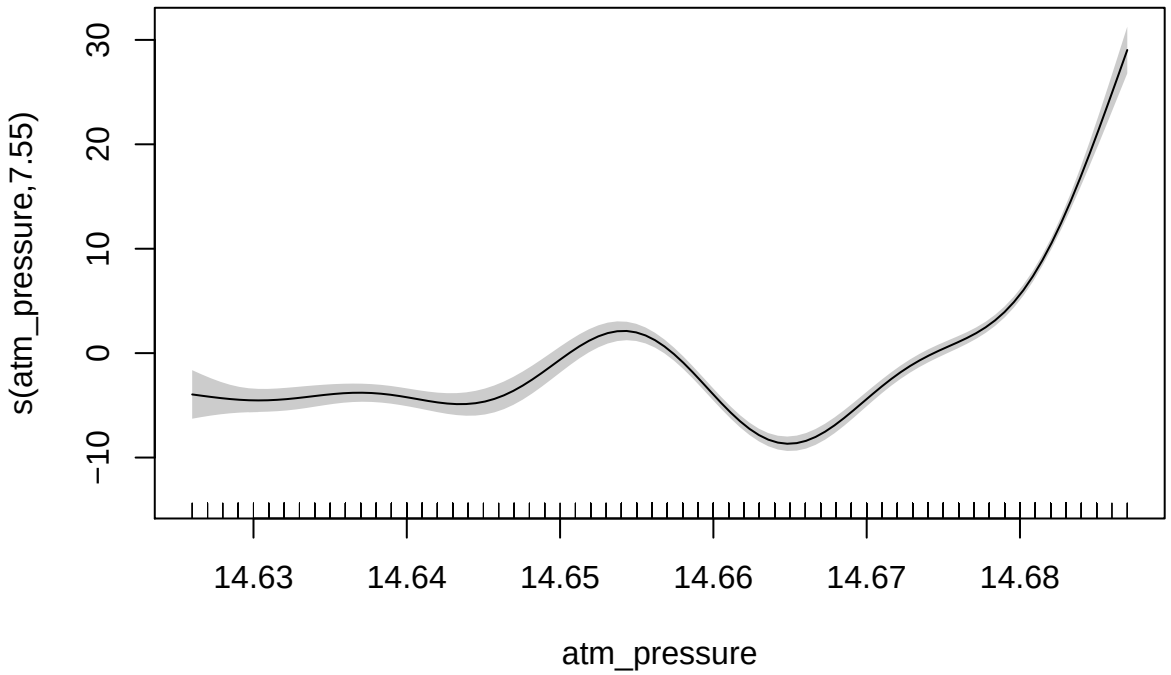
```
semi_mod <- mgcv::gam(freq ~ s(int_pressure) + s(atm_pressure), data=mri)
plot(semi_mod, shade=T)
```






```
semi_mod2 <- mgcv::gam(diff(freq) ~ s(int_pressure[2:length(int_pressure)]) +
                         s(atm_pressure[2:length(atm_pressure)])
                       , data=mri)
plot(semi_mod, shade=T)
```





The graphics are showing the individual effects of each of the predictors on the outcome of *frequency*, controlling for other predictors. The bottom of each graphic has a "rug" that will show you where the data points lie.

In both cases, the effect of *internal_pressure* is unclear - most of the points are centered around a few small values. Atmospheric pressure seems to tell a more interesting story - as we increase the value of it, its effects on the predictor increase.